

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
22 July 2004 (22.07.2004)

PCT

(10) International Publication Number  
**WO 2004/061124 A2**

- (51) International Patent Classification<sup>7</sup>: **C12Q** **BATES, Stephen** [US/US]; 1311 Torrey Street, Davis, CA 95616 (US).
- (21) International Application Number: PCT/US2003/041761 (74) Agent: **HAILE, Lisa, A.**; Gray Cary Ware & Freidenrich LLP, Suite 1100, 4365 Executive Drive, San Diego, CA 92121-2133 (US).
- (22) International Filing Date: 31 December 2003 (31.12.2003) (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/437,482 31 December 2002 (31.12.2002) US (84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application: US 60/437,482 (CIP) Filed on 31 December 2002 (31.12.2002)
- (71) Applicant (*for all designated States except US*): **MMI GENOMICS, INC.** [US/US]; 1756 Picasso Avenue, Davis, CA 95616 (US).
- (72) Inventors; and (75) Inventors/Applicants (*for US only*): **DeNISE, Sue, K.** [US/US]; 412 Alvarado Avenue, Davis, CA 95616 (US). **CHARTERIS, Paul** [NZ/US]; 603 G Street, Davis, CA 95616 (US). **ROSENFELD, David** [US/US]; 3431 David Way, Sacramento, CA 95820 (US). **HOLM, Tom** [US/US]; 1843 E. Hillcrest Avenue, Salt Lake City, UT 84106 (US).
- Published:**
- without international search report and to be republished upon receipt of that report
  - with sequence listing part of description published separately in electronic form and available upon request from the International Bureau
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: COMPOSITIONS, METHODS, AND SYSTEMS FOR INFERRING BOVINE BREED

(57) Abstract: Provided herein are methods to discover and use single nucleotide polymorphisms (SNP) for identifying breed, or line and breed, or line composition of a bovine subject. The present invention further provides specific nucleic acid sequences, SNPs, and SNP patterns that can be used for identifying breed or breed combinations for Angus, Holstein, Limousin, Brahman, Hereford, Simmental, Gelbvieh, Charolais and Beefmaster breeds. These patterns can be utilized to manage animals in a feedlot to obtain optimum performance based on known characteristics of specific breeds and identify animals for breeding in selection programs. In another aspect, these patterns can be used to ensure labeling on breed specific branded products.

WO 2004/061124 A2

## COMPOSITIONS, METHODS, AND SYSTEMS FOR INFERRING BOVINE BREED

### CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of priority under 35 U.S.C. § 119(e) of U.S. Serial No. 60/437,482, filed December 31, 2002, the entire content of which is incorporated herein by reference.

### BACKGROUND OF THE INVENTION

#### FIELD OF THE INVENTION

[0002] The invention relates generally to genetic markers and more specifically to polymorphisms associated with bovine breed.

#### BACKGROUND INFORMATION

[0003] Classification of individual animals in a livestock population has often relied on *a priori* groupings of individual animals on the basis of parentage and registration with a Breed Association. If these criteria are not known or not available, animals can be classified as a member of a breed or combination of breeds based on phenotype or geographic location. For example, a bovine animal with a black coat color is frequently classified as an Angus – a breed distinguished by having a black coat color. Further, a bovine animal with a pronounced hump over the shoulder region, pendulous ears and loose skin on dewlap and throat is classified as Brahman. These phenotypes such as coat color, ear and hump appearance are readily observable by livestock producers and frequently used for the basis of breed classification.

[0004] Two possible options for classifying an individual bovine animal into a population are:

- 1) Assign an animal to a population based on known or assumed parentage, phenotypic appearance or trait value for some phenotype, or
- 2) From a set of predefined populations (such as breeds), sample DNA from a number of members of each population to estimate allele frequencies in each population. Using the allele frequencies, it is possible to compute the likelihood a

given genotype originated in each population and individuals can be assigned to population on the basis of these likelihoods (Pritchard, J.K., et al., *Genetics* 155: 945-959 (2000)).

[0005] Both strategies (above) rely on defining a set of populations. A classification based on phenotype or geographic locality may not accurately describe the genetic structure of a population if similar phenotypes can arise despite differences in genotype (Rosenberg, N.A., et al., *Genetics*. 159: 699-713 (2001)). Coat color in cattle is determined by one or a few loci that are inherited in a well-known manner (Olson, T.A. 1999. Genetics of colour variation. In: Fries, R. and Ruvinsky, A. (eds.) *The Genetics of Cattle*- CABI Publishing, Wallingford.). It is relatively easy for livestock breeders to introgress new genes for coat color into a population (such as a breed) resulting in phenotypes that were not previously present.

[0006] Some beef marketing systems rely on accurate determination of breed of animal. Of the 41 brands of beef certified by the USDA, 33 name a breed and of these, 30 name Angus, two name Hereford and one names Piedmontese (Smith, G.C., available on the internet at [ansci.colostate.edu/ran/beef/smith7.pdf](http://ansci.colostate.edu/ran/beef/smith7.pdf)). To date, the only methods available to qualify animals for these systems are known or assumed parentage or phenotypic appearance. There is an opportunity to improve accuracy of individual animal qualification using the allele frequencies to compute the likelihood that a given genotype originated in specific breed population.

[0007] It is critical to know the breed of cattle entered into and qualifying from branded beef programs when those branded beef programs include breed specifications. In particular, knowledge of breed composition is important for the following:

- 1) to verify claims for breed type or breed composition associated with breed-specific marketing programs. Verification of claims for breed type or breed composition has not been possible because no available technology could classify a bovine animal to a particular population or infer the breed composition of an individual animal. Currently, the only bovines accepted by breed are those within the seedstock sector where the records of individual animals are maintained by Breed

Associations or in commercial populations where cattle are recorded in a performance registry – usually by a Breed Association. These total breed-identified cattle likely comprise some five percent of the national beef cattle population. Therefore, standardization of claims for breed or breed type is very loose and limited to a small fraction of animals contributing to the beef supply; and

- 2) to implement management strategies for the feeding of cattle to optimize the pre-harvest growth and development and post-harvest fabrication of their associated beef products. Efficiencies in production from feeding, sorting, managing and marketing cattle can be obtained by identifying and matching breed-specific phenotypic data for feedlot, carcass and meat quality traits. Where breeds cannot be distinguished on the merit of parentage and simply measured phenotypic differences, a purely genetic analysis provides the most suitable test to assign animals to a population (Rosenberg, N.A., et al., *Genetics* 159: 699-713 (2001)). Only if a correspondence between genotype and known parentage classification is established can these characteristics also serve as appropriate classification tools.

[0008] Accordingly, there remains a need for methods and compositions that provide information regarding bovine breed.

#### **SUMMARY OF THE INVENTION**

[0009] The present invention is based, in part, on the discovery of bovine single nucleotide polymorphism (SNP) markers that are associated with, and predictive of, bovine breeds including, but not limited to, Angus, Brahman, Charolais, Hereford, Simmental, Gelbvieh, Limousin, and Beefmaster breeds. Accordingly, the present invention provides methods to discover and use single nucleotide polymorphisms (SNP) for identifying breed, or line and breed, or line composition of a bovine subject. The present invention further provides specific nucleic acid sequences, SNPs, and SNP patterns that can be used for identifying breed or breed combinations for Angus, Limousin, Brahman, Hereford, Simmental, Gelbvieh, Charolais and Beefmaster breeds. These patterns can be utilized to manage animals in a feedlot to obtain optimum

performance based on known characteristics of specific breeds. In another embodiment, cattle can be marketed prior to harvest based on their breed characteristics for meat production. In another aspect, branded products based on breed designation can certify their label based on specific breed markers.

[0010] Accordingly, in one embodiment the present invention provides a method to infer breed of a bovine subject from a nucleic acid sample of the bovine subject, that includes identifying in the nucleic acid sample, at least one nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with a breed, thereby inferring the breed of the bovine subject. A SNP is associated with a breed when at least one nucleotide occurrence of the SNP occurs more frequently in subjects of a particular breed than other breeds in a statistically significant manner, for example with greater than 80%, 85%, 90%, 95%, or 99% confidence. Therefore, in certain aspects, the methods include identifying whether the nucleotide occurrence is a bovine SNP allele identified herein as associated with bovine breed. In certain aspects, the identified breed includes, but is not limited to, Angus, Limousin, Brahman, Simmental, Hereford, Gelbvieh or Charolais. Furthermore, in certain aspects, at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Angus cattle is identified.

[0011] In another embodiment, the present invention provides a method for determining a nucleotide occurrence of a single nucleotide polymorphism (SNP) in a bovine sample, that includes contacting a bovine polynucleotide in the sample with an oligonucleotide that binds to a target region of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, and determining the nucleotide occurrence of a single nucleotide polymorphism (SNP) corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895. The determination typically includes analyzing binding of the oligonucleotide, or detecting an amplification product generated using the oligonucleotide, thereby determining the nucleotide occurrence of the SNP.

[0012] In another embodiment, the present invention provides an isolated polynucleotide that includes a fragment of at least 20 contiguous nucleotides, a polynucleotide at least 90% identical to the fragment of 20 contiguous nucleotides, or a complement thereof, wherein the isolated polynucleotide includes a nucleotide occurrence of a single nucleotide polymorphism (SNP) associated with breed, wherein the SNP corresponds to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

[0013] In yet another embodiment, the present invention provides an isolated oligonucleotide that includes 10 nucleotides, that selectively binds to a target polynucleotide of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein a terminal nucleotide of the isolated oligonucleotide binds to position 299, 300, or 301 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

#### **DETAILED DESCRIPTION OF THE INVENTION**

[0014] The specification hereby incorporates by reference in their entirety, the files contained on the two compact discs filed herewith. Two copies of each of the two compact discs are filed herewith. The first compact disc includes a file called "mmi1150wo Table 5.doc," created December 31, 2003, which is 4770 kilobytes in size, and a file called "mmi1150wo Table 6.doc," created December 31, 2003, which is 85875 kilobytes in size. The Second disc includes a sequence listing which is included in a file called "MMI1150WO SEQUENCE LISTING.txt," created December 31, 2003, which is 88096 kilobytes in size.

[0015] The present invention is based in part on the discovery of single nucleotide polymorphisms (SNPs) that can be used to infer breed of a bovine subject. Accordingly, provided herein is a method for inferring the breed of a bovine subject from a nucleic acid sample of the bovine subject, by identifying in the sample, a nucleotide occurrence for at least one single nucleotide polymorphism (SNP), wherein the nucleotide occurrence is associated with the breed. In certain aspects, the SNP corresponds to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

[0016] Using the teachings herein, SNPs associated with any breed of cattle can be identified. Therefore, methods of the present invention for inferring breed of a bovine

subject, can be used to infer the breed of any bovine subject. For example, the methods can be used to infer a breed including, but not limited to, Angus, Limousin, Brahman, Simmental, Hereford, Holstein, Gelbvieh or Charolais cattle. In certain aspects, the methods are used to infer an Angus from a non-Angus breed. Furthermore, the methods of the present invention can be used to assign a breed or breeds to an individual animal with a specific probability. Typically, an identified nucleotide occurrence is compared to multiple known SNP alleles associated with multiple breeds, for example the breed associated alleles identified herein in Tables 5 and 3A, to infer a breed for a subject from multiple possible breeds.

[0017] The Example provided herein illustrates the use of whole genome shotgun sequencing and genotyping analysis to identify SNPs that can be used to infer breed of a bovine subject. For this analysis, genomic DNA libraries were constructed from six breeds of cattle (Angus, Limousin, Brahman, Simmental, Holstein and Charolais). The libraries were a mixture of 2.5, and 10 kilobase inserts. Libraries from four of the cattle breeds (Angus, Brahman, Simmental and Limousin) were sequenced, using the whole genome shotgun sequencing method developed by Celera Genomics (Venter, J.C. *et al.* (2001) *Science* 291:1304), to a depth of coverage sufficient to generate putative SNPs. The distribution of sequence fragments obtained from the cattle breeds was 32% Angus, 35% Limousin, 33% Brahman, and less than 1% Simmental.

[0018] Allele frequencies within breed were determined using simple counting methods. Four thousand eight hundred and seventy seven breed specific markers (See SEQ ID NOS:1 to 4868 (Tables 5 and 6, included on a compact disc filed herewith) and SEQ ID NOS:64887 to 64895 (Tables 3A and 3B)) were identified by analysis of over 19,000 markers tested. Details of the SNP markers that were identified as informative for breed, are listed in Tables 5 and 6 (filed herewith on a compact disc) and Tables 3A and 3B. Tables 3A and 5 provide the names of the identified SNP markers, breed specific alleles, their associated breed, allele frequencies, and breeds for which the alleles, and therefore SNPs, are most significantly associated (referred to as "major breeds" in the table). A SNP was identified as being associated with a breed when the frequency of at least one allele was significantly greater for at least one breed than for other breeds. Tables 3B and 6 provide primer sequences (See "Forward," "Reverse," and "Extension")

that were used to amplify a region that includes the SNP, and amplicon sequences that indicate the nucleotide occurrences for the SNP that were identified in brackets within the sequence.

[0019] Accordingly, in one aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Angus cattle. In another aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Brahman cattle. In another aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Limousin cattle. In yet another aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Simmental cattle. In still another aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Hereford cattle. In another aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Charolais cattle. In another aspect, the method identifies at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Gelbvieh cattle.

[0020] In certain aspects, the method further includes marketing the bovine subject prior to harvest based on the inferred breed. As indicated herein, some beef marketing systems rely on accurate determination of a breed of an animal. For example, in certain aspects the inferred breed is used to assign beef of a bovine subject to a USDA certified brand before the subject is harvested.

[0021] In another aspect, the method further includes managing the bovine subject in the feedlot to obtain improved performance based on known characteristics of the inferred breed for the subject. In fact, the SNPs and methods of the present invention can be used to identify the percentage of each breed comprising an individual animal. In this aspect, for example, animals arriving to the feedlot can have a DNA sample taken at processing. The sample can be screened against breed specific DNA markers to determine the percentage of breed per animal based on the breed specific marker panel and the degree of accuracy required. These data can be used to manage feedlot cattle for specific growth



and development traits. In one aspect, cattle identified from the exotic breed type (Charolais, Gelbvieh, Limousin, and Simmental) have high growth rate and large harvest sizes. These cattle can be managed to maximize growth rate and lean meat yield. Cattle identified as English type (Angus and Hereford) produce high quality meat products and can be marketed into the high quality yield grade.

[0022] In other examples, methods of the present invention further include identifying a nucleotide sequence of a hypermutable sequence in the sample, and inferring breed based on at least one SNP nucleotide occurrence and the nucleotide sequence of the hypermutable sequence. Hypermutable sequences include, for example, microsatellite nucleic acid sequences

[0023] In another aspect, a method of the present invention further includes branding or marketing a product of the bovine subject based on the inferred breed, including marketing the product under a trademark specific for breed. Breed specified products can certify that the product is in fact from the breed labeled, thereby increasing consumer confidence in a product and increasing the value of a product.

[0024] The method can include a determination of the nucleotide occurrence of at least 2 SNPs. At least 2 SNPs can form all or a portion of a haplotype, wherein the method identifies a haplotype allele that is associated with a specific breed. Furthermore, the method can include identifying a diploid pair of haplotype alleles.

[0025] In another embodiment, the present invention provides a method for sorting one or more bovine subjects, that includes inferring a breed for a first bovine subject from a nucleic acid sample of the first bovine subject, by identifying a nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of at least one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with breed. The first bovine subject is sorted based on the inferred breed. The method can then be repeated for additional subjects, thereby sorting bovine subjects. The bovine subjects can be sorted, for example, based on whether they are Angus, Limousin, Brahman, Simmental, Hereford, Gelbvieh or Charolais cattle.

[0026] In another embodiment, the present invention provides a method of providing labeling accuracy for breed identified meat products, that includes inferring breed of a

bovine candidate for use in branded meat products from a nucleic acid sample of the bovine candidate by a method comprising identifying the nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of at least one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with breed; and determining whether to brand the harvested product based on the inferred breed. This method provides quality assurance guarantees of the label brand. In certain aspects, for example, the inferred breed is Angus, Limousin, Brahman, Simmental, Hereford, Gelbvieh or Charolais.

[0027] In yet another embodiment, the present invention provides a method for breeding a bovine subject, that includes inferring breed of a bovine candidate for use in breeding programs from a nucleic acid sample of the bovine candidate by a method that includes identifying the nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of at least one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with breed. A determination is made on whether to select the individual for use in breeding programs based on the inferred breed, thereby breeding the bovine subject.

[0028] The selection is implemented, in certain aspects, at the elite or breeding nucleus level or at the multiplier or foundation animal level. In another embodiment, the present invention provides a bovine subject resulting from the selection and breeding.

[0029] In another embodiment, the present invention provides a method for identifying a bovine single nucleotide polymorphism (SNP) informative of breed, that includes performing whole genome shotgun sequencing of a bovine genome, and genotyping at least two bovine subjects from at least two breeds, thereby identifying the bovine single nucleotide polymorphisms informative of breed. The Example provided herein, illustrates the use of this method to identify breed SNPs.

[0030] As used herein, the term "at least one", when used in reference to a gene, SNP, haplotype, or the like, means 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc., up to and including all of the haplotype alleles, genes, haplotypes, and/or SNPs of the bovine genome. Reference to "at least a second" gene, SNP, haplotype or the like, means two or more, i.e., 2, 3, 4, 5, 6, 7, 8, 9, 10, etc., bovine genes, SNPs, haplotypes, or the like.

[0031] Polymorphisms are allelic variants that occur in a population that can be a single nucleotide difference present at a locus, or can be an insertion or deletion of one, a few or many consecutive nucleotides. As such, a single nucleotide polymorphism (SNP) is characterized by the presence in a population of one or two, three or four nucleotides (i.e., adenosine, cytosine, guanosine or thymidine), typically less than all four nucleotides, at a particular locus in a genome such as the human genome. It will be recognized that, while the methods of the invention are exemplified primarily by the detection of SNPs, the disclosed methods or others known in the art similarly can be used to identify other types of bovine polymorphisms, which typically involve more than one nucleotide. A SNP is associated with a breed when at least one nucleotide occurrence of the SNP occurs more frequently in subjects of a particular breed in a statistically significant manner, for example with greater than 80%, 85%, 90%, 95%, or 99% confidence. A bovine "SNP allele" is a nucleotide occurrence of a SNP within a population of bovine animals.

[0032] In another embodiment, the present invention provides an isolated polynucleotide that includes a fragment of at least 20 contiguous nucleotides of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, a polynucleotide at least 90% identical to the 20 contiguous nucleotide fragment, or a complement thereof, wherein the isolated polynucleotide includes a nucleotide occurrence of a single nucleotide polymorphism (SNP) that corresponds to position 300 of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

[0033] In certain aspects, the isolated polynucleotide, for example, includes a fragment of at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, 250, 500, or 600 contiguous nucleotides of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895. In another aspect, the isolated polynucleotide is at least 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, or 99.5% identical to SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, for example. Typically, the isolated nucleotide includes a region that is contiguous with a region of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895 that includes position 300. In certain aspects, the isolated polynucleotide consists of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895 in its entirety. In other aspects, the isolated polynucleotide consists of any one of SEQ ID NOS:4869 to 19472 or SEQ ID NOS:64896 to 64922.

[0034] The polynucleotide or an oligonucleotide of the invention can further include a detectable label. For example, the detectable label can be associated with the polynucleotide at a position corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895. As discussed in more detail herein, the labeled polynucleotide can be generated, for example, during a microsequencing reaction, such as SNP-IT™ reaction.

[0035] Detectable labeling of a polynucleotide or oligonucleotide is well known in the art. Particular non-limiting examples of detectable labels include chemiluminescent labels, fluorescent labels, radiolabels, enzymes, haptens, or even unique oligonucleotide sequences.

[0036] In another embodiment, the present invention provides an isolated vector that includes a polynucleotide disclosed hereinabove. The term "vector" refers to a plasmid, virus or other vehicle known in the art that has been manipulated by insertion or incorporation of a nucleic acid sequence.

[0037] Methods that are well known in the art can be used to construct vectors, including *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* recombination/genetic techniques (See, for example, the techniques described in Maniatis et al. 1989 Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y., incorporated herein in its entirety by reference).

[0038] In another aspect, the present invention provides an isolated cell that includes the vector. The cell can be prokaryotic or eukaryotic. Techniques for incorporated vectors into prokaryotic and eukaryotic cells are well known in the art. In certain aspects, the cells are bovine cells. In other aspects, the cells are bacterial cells. In still other aspects, the cells are human cells.

[0039] In another aspect, the present invention provides a primer pair that binds to a first target region and a second target region of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the first primer of the primer pair and a second primer of the primer pair are at least 10 nucleotides in length and bind opposite strands of the target region located within 3000 nucleotides of a position corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, and prime

polynucleotide synthesis from the target region in opposite directions across position 300. In another embodiment, provided herein is a primer pair that binds to a first target region and a second target region of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein a first primer of the primer pair and a second primer of the primer pair are at least 10 nucleotides in length, bind opposite strands of the target region, and prime polynucleotide synthesis from the target region in opposite directions across position 300 of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

[0040] In another embodiment, the present invention provides an isolated oligonucleotide that selectively binds to a target polynucleotide that includes at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 300, 500, or 600 nucleotides SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, or a complement thereof, wherein the terminal nucleotide corresponds to position 299, 300, or 301. In another embodiment, the present invention provides an isolated oligonucleotide that includes 10 nucleotides, which selectively binds to a target polynucleotide of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein a terminal nucleotide of the isolated oligonucleotide binds to position 298, 299, 300, 301, or 302 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895. The oligonucleotide can be, for example, 10, 15, 20, 25, 50, or 100 nucleotides in length. In certain aspects, the terminal nucleotide binds to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

[0041] In another embodiment, the present invention provides an isolated oligonucleotide pair effective for determining a nucleotide occurrence at a single nucleotide polymorphism (SNP) corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein each isolated oligonucleotide comprises at least 5 nucleotides from SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895 and wherein the terminal nucleotide of each oligonucleotide pair is complementary to a different nucleotide at position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895 or a complement thereof. In certain aspects, the specific binding pair member is a substrate for a primer extension reaction.

[0042] As used herein, "about" means within ten percent of a value. For example, "about 100" would mean a value between 90 and 110.

[0043] The term "haplotypes" as used herein refers to groupings of two or more SNPs that are physically present on the same chromosome which tend to be inherited together except when recombination occurs. The haplotype provides information regarding an allele of the gene, regulatory regions or other genetic sequences affecting a trait. The linkage disequilibrium and, thus, association of a SNP or a haplotype allele(s) and a bovine breed can be strong enough to be detected using simple genetic approaches, or can require more sophisticated statistical approaches to be identified.

[0044] Numerous methods for identifying haplotype alleles in nucleic acid samples are known in the art. In general, nucleic acid occurrences for the individual SNPs are determined and then combined to identify haplotype alleles. There are several algorithms for haplotype reconstruction based on pedigree analysis. These are the Maximum Likelihood methods ((Excofier, L., and Slatkin, M., *Mol. Biol. Evol.* 12: 921-927 (1995)), the parsimony method created by Clark, A.G., *Mol. Biol. Evol.* 7: 111-122 (1990) and the phase reconstruction method of Stephens, M., et al., *Am. J. Hum. Genet.* 68:978-989, 2001, which is incorporated herein by reference). These methods can be applied to the data generated, regarding individual nucleotide occurrences in SNP markers of the subject, in order to determine alleles for each haplotype in a subject's genotype. Alternatively, haplotypes can also be determined directly, for each pair of sites, by allele-specific PCR (Clark, A.G. et al., *Am. J. Hum. Genet.* 63: 595-612 (1998)).

[0045] As used herein, the term "infer" or "inferring", when used in reference to a breed, means drawing a conclusion about a breed using a process of analyzing individually or in combination, nucleotide occurrence(s) of one or more SNP(s), which can be part of one or more haplotypes, in a nucleic acid sample of the subject, and comparing the individual or combination of nucleotide occurrence(s) of the SNP(s) to known relationships of nucleotide occurrence(s) of the SNP(s) and the breed. As disclosed herein, the nucleotide occurrence(s) can be identified directly by examining nucleic acid molecules, or indirectly by examining a polypeptide encoded by a particular gene where the polymorphism is associated with an amino acid change in the encoded polypeptide.

[0046] Relationships between nucleotide occurrences of one or more SNPs or haplotypes and a breed can be identified using known statistical methods. A statistical analysis result which shows an association of one or more SNPs or haplotypes with a breed with at least 80%, 85%, 90%, 95%, or 99% confidence, or alternatively a probability of insignificance less than 0.05, can be used to identify SNPs and haplotypes. These statistical tools may test for significance related to a null hypothesis that an on-test SNP allele or haplotype allele is not significantly different between groups with different traits. If the significance of this difference is low, it suggests the allele is not related to a breed.

[0047] In diploid organisms such as bovines, somatic cells, which are diploid, include two alleles for each single-locus haplotype. As such, in some cases, the two alleles of a haplotype are referred to herein as a genotype or as a diploid pair, and the analysis of somatic cells, typically identifies the alleles for each copy of the haplotype. Methods of the present invention can include identifying a diploid pair of haplotype alleles. These alleles can be identical (homozygous) or can be different (heterozygous). Haplotypes that extend over multiple loci on the same chromosome include up to 2 to the Nth power alleles where N is the number of loci. It is beneficial to express polymorphisms in terms of multi-locus (i.e. multi SNP) haplotypes because haplotypes offer enhanced statistical power for genetic association studies. Multi-locus haplotypes can be precisely determined from diploid pairs when the diploid pairs include 0 or 1 heterozygous pairs, and N or N-1 homozygous pairs. When multi-locus haplotypes cannot be precisely determined, they can sometimes be inferred by statistical methods. Methods of the invention can include identifying multi-locus haplotypes, either precisely determined, or inferred.

[0048] A sample useful for practicing a method of the invention can be any biological sample of a subject, typically a bovine subject, that contains nucleic acid molecules, including portions of the gene sequences to be examined, or corresponding encoded polypeptides, depending on the particular method. As such, the sample can be a cell, tissue or organ sample, or can be a sample of a biological material such as blood, milk, semen, saliva, hair, tissue, and the like. A nucleic acid sample useful for practicing a method of the invention can be deoxyribonucleic (DNA) acid or ribonucleic acids (RNA). The nucleic acid sample generally is a deoxyribonucleic acid sample, particularly

genomic DNA or an amplification product thereof. However, where heteronuclear ribonucleic acid, which includes unspliced mRNA precursor RNA molecules and non-coding regulatory molecules such as RNA, is available, a cDNA or amplification product thereof can be used.

[0049] Where each of the SNPs of the haplotype is present in a coding region of a gene(s), the nucleic acid sample can be DNA or RNA, or products derived therefrom, for example, amplification products. Furthermore, while the methods of the invention generally are exemplified with respect to a nucleic acid sample, it will be recognized that particular haplotype alleles can be in coding regions of a gene and can result in polypeptides containing different amino acids at the positions corresponding to the SNPs due to non-degenerate codon changes. As such, in another aspect, the methods of the invention can be practiced using a sample containing polypeptides of the subject.

[0050] In one embodiment, DNA samples are collected and stored in a retrievable barcode system, either automated or manual, that ties to a database. Collection practices include systems for collecting tissue, hair, mouth cells or blood samples from individual animals at the same time that ear tags, electronic identification or other devices are attached or implanted into the animal. All identities of animals can be automatically uploaded into a primary database. Tissue collection devices can be integrated into the tool used for placing the ear tag. Body fluid samples can be collected and stored on a membrane bound system.

[0051] The sample is then analyzed on the premises or sent to a laboratory where a medium to high-throughput genotyping system is used to analyze the sample.

[0052] The subject of the present invention can be any bovine subject, for example a bull, a cow, a calf, a steer, or a heifer or any bovine embryo or tissue.

[0053] In another aspect, the present invention provides a system for determining the nucleotide occurrences in a population of bovine single nucleotide polymorphisms (SNPs). The system typically includes a hybridization medium and/or substrate that includes at least two oligonucleotides of the present invention, or oligonucleotides used in the methods of the present invention. The hybridization medium and/or substrate are used to determine the nucleotide occurrence of bovine SNPs that are associated with



breed. Accordingly, the oligonucleotides are used to determine the nucleotide occurrence of bovine SNPs that are associated with a breed. The determination can be made by selecting oligonucleotides that bind at or near a genomic location of each SNP of the series of bovine SNPs. The system of the present invention typically includes a reagent handling mechanism that can be used to apply a reagent, typically a liquid, to the solid support. The binding of an oligonucleotide of the series of oligonucleotides to a polynucleotide isolated from a genome can be affected by the nucleotide occurrence of the SNP. The system can include a mechanism effective for moving a solid support and a detection mechanism. The detection method detects binding or tagging of the oligonucleotides.

[0054] Accordingly, in another embodiment, the present invention provides a method for determining a nucleotide occurrence of a single nucleotide polymorphism (SNP) in a bovine sample, that includes contacting a bovine polynucleotide in the sample with an oligonucleotide that binds to a target region of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895. The nucleotide occurrence of a single nucleotide polymorphism (SNP) corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, is then determined, wherein the determination comprises analyzing binding of the oligonucleotide, or detecting an amplification product generated using the oligonucleotide, thereby determining the nucleotide occurrence of the SNP. In certain aspects, the oligonucleotide binds to a target sequence that includes one of the SNPs, and the nucleotide occurrence is determined based on the binding of the oligonucleotide to the target sequence.

[0055] In another aspect, the bovine polynucleotide is contacted with a pair of oligonucleotides that constitute a primer pair, and the nucleotide occurrence is determined using an amplification product generated using the primer pair. For example, at least one primer of the primer pair can be one of SEQ ID NOS:Forward1 to Reverse4868. Furthermore, the primer pair, in certain aspects, is any of the forward and reverse primer pairs listed in Table 6 (Appendix 2).

[0056] In certain aspects, the terminal nucleotide of the oligonucleotide binds to the SNP. In these aspects, the method can include detecting an extension product generated using

the oligonucleotide as a primer. The oligonucleotide, in certain aspects, is any one of SEQ ID NO:4879 to 9736 or SEQ ID NO:64914 to 64922.

[0057] In other aspects, the terminal nucleotide of each oligonucleotide of a pair of oligonucleotides is complementary to a different nucleotide at position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, or a complement thereof. In another example, the polynucleotide can be contacted with a pair of oligonucleotides each comprising a different detectable label.

[0058] Medium to high-throughput systems for analyzing SNPs, known in the art such as the SNPStream® UHT Genotyping System (Beckman/Coulter, Fullerton, CA) (Boyce-Jacino and Goelet Patents), the Mass Array™ system (Sequenom, San Diego, CA) (Storm, N. et al. (2002) *Methods in Molecular Biology* 212: 241-262.), the BeadArray™ SNP genotyping system available from Illumina (San Diego, CA) (Oliphant, A., et al. (June 2002) (supplement to *Biotechniques*), and TaqMan™ (Applied Biosystems, Foster City, CA) can be used with the present invention. However, the present invention provides a medium to high-throughput system that is designed to detect nucleotide occurrences of bovine SNPs, or a series of bovine SNPs that can make up a series of haplotypes. Therefore, as indicated above the system includes a solid support or other method to which a series of oligonucleotides can be associated that are used to determine a nucleotide occurrence of a SNP for a series of bovine SNPs that are associated with a trait. The system can further include a detection mechanism for detecting binding of the series of oligonucleotides to the series of SNPs. Such detection mechanisms are known in the art.

[0059] The system can be a microfluidic device. Numerous microfluidic devices are known that include solid supports with microchannels (See e.g., U.S. Pat. Nos. 5,304,487, 5,110,745, 5,681,484, and 5,593,838).

[0060] The SNP detection systems of the present invention are designed to determine nucleotide occurrences of one SNP or a series of SNPs. The systems can determine nucleotide occurrences of an entire genome-wide high-density SNP map.

[0061] Numerous methods are known in the art for determining the nucleotide occurrence for a particular SNP in a sample. Such methods can utilize one or more oligonucleotide

probes or primers, including, for example, an amplification primer pair that selectively hybridizes to a target polynucleotide, which corresponds to one or more bovine SNP positions. Oligonucleotide probes useful in practicing a method of the invention can include, for example, an oligonucleotide that is complementary to and spans a portion of the target polynucleotide, including the position of the SNP, wherein the presence of a specific nucleotide at the position (i.e., the SNP) is detected by the presence or absence of selective hybridization of the probe. Such a method can further include contacting the target polynucleotide and hybridized oligonucleotide with an endonuclease, and detecting the presence or absence of a cleavage product of the probe, depending on whether the nucleotide occurrence at the SNP site is complementary to the corresponding nucleotide of the probe. These oligonucleotides and probes are another embodiment of the present invention.

[0062] An oligonucleotide ligation assay (Grossman, P.D. et al. (1994) *Nucleic Acids Research* 22:4527-4534) also can be used to identify a nucleotide occurrence at a polymorphic position, wherein a pair of probes that selectively hybridize upstream and adjacent to and downstream and adjacent to the site of the SNP, and wherein one of the probes includes a terminal nucleotide complementary to a nucleotide occurrence of the SNP. Where the terminal nucleotide of the probe is complementary to the nucleotide occurrence, selective hybridization includes the terminal nucleotide such that, in the presence of a ligase, the upstream and downstream oligonucleotides are ligated. As such, the presence or absence of a ligation product is indicative of the nucleotide occurrence at the SNP site. An example of this type of assay is the SNPLex System (Applied Biosystems, Foster City, CA).

[0063] An oligonucleotide also can be useful as a primer, for example, for a primer extension reaction, wherein the product (or absence of a product) of the extension reaction is indicative of the nucleotide occurrence. In addition, a primer pair useful for amplifying a portion of the target polynucleotide including the SNP site can be useful, wherein the amplification product is examined to determine the nucleotide occurrence at the SNP site. Particularly useful methods include those that are readily adaptable to a high throughput format, to a multiplex format, or to both. The primer extension or amplification product can be detected directly or indirectly and/or can be sequenced using

various methods known in the art. Amplification products which span a SNP locus can be sequenced using traditional sequence methodologies (e.g., the "dideoxy-mediated chain termination method," also known as the "Sanger Method" (Sanger, F., et al., J. Molec. Biol. 94:441 (1975); Prober et al. Science 238:336-340 (1987)) and the "chemical degradation method," also known as the "Maxam-Gilbert method" (Maxam, A. M., et al., Proc. Natl. Acad. Sci. (U.S.A.) 74:560 (1977)), both references herein incorporated by reference) to determine the nucleotide occurrence at the SNP locus.

[0064] Methods of the invention can identify nucleotide occurrences at SNPs using genome-wide sequencing or "microsequencing" methods. Whole-genome sequencing of individuals identifies all SNP genotypes in a single analysis. Microsequencing methods determine the identity of only a single nucleotide at a "predetermined" site. Such methods have particular utility in determining the presence and identity of polymorphisms in a target polynucleotide. Such microsequencing methods, as well as other methods for determining the nucleotide occurrence at a SNP locus are discussed in Boyce-Jacino, et al., U.S. Pat. No. 6,294,336, incorporated herein by reference, and summarized herein.

[0065] Microsequencing methods include the Genetic Bit™ Analysis method disclosed by Goelet, P. et al. (WO 92/15712, herein incorporated by reference). Additional, primer-guided, nucleotide incorporation procedures for assaying polymorphic sites in DNA have also been described (Kornher, J. S. et al, Nucleic Acids Res. 17:7779-7784 (1989); Sokolov, B. P., Nucleic Acids Res. 18:3671 (1990); Syvanen, A. -C., et al., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et al., Proc. Natl. Acad. Sci. (U.S.A.) 88:1143-1147 (1991); Prezant, T. R. et al, Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et al., GATA 9:107-112 (1992); Nyren, P. et al., Anal. Biochem. 208:171-175 (1993); and Wallace, WO89/10414). These methods differ from Genetic Bit™ Analysis in that they all rely on the incorporation of labeled deoxynucleotides to discriminate between bases at a polymorphic site. In such a format, since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A. -C., et al. *Amer. J. Hum. Genet.* (1993) 52:46-59 Other formats for microsequencing include

Pyrosequencing (Pyrosequencing AB, Uppsala, Sweden, Alderborn et al (2000) *Genome Res.* 10:1249-1258).

[0066] Alternative microsequencing methods have been provided by Mundy, C.R. (U.S. Pat. No. 4,656,127) and Cohen, D. et al (French Patent 2,650,840; PCT Appln. No. WO91/02087), which discuss a solution-based method for determining the identity of the nucleotide of a polymorphic site. As in the Mundy method of U.S. Pat. No. 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3'-to a polymorphic site.

[0067] In response to the difficulties encountered in employing gel electrophoresis to analyze sequences, alternative methods for microsequencing have been developed. Macevicz (U.S. Pat. No. 5,002,867), for example, describes a method for determining nucleic acid sequence via hybridization with multiple mixtures of oligonucleotide probes. In accordance with such method, the sequence of a target polynucleotide is determined by permitting the target to sequentially hybridize with sets of probes having an invariant nucleotide at one position, and variant nucleotides at other positions. The Macevicz method determines the nucleotide sequence of the target by hybridizing the target with a set of probes, and then determining the number of sites that at least one member of the set is capable of hybridizing to the target (i.e., the number of "matches"). This procedure is repeated until each member of a set of probes has been tested.

[0068] Boyce-Jacino, et al., U.S. Pat. No. 6,294,336 provides a solid phase sequencing method for determining the sequence of nucleic acid molecules (either DNA or RNA) by utilizing a primer that selectively binds a polynucleotide target at a site wherein the SNP is the most 3' nucleotide selectively bound to the target.

[0069] The occurrence of a SNP can be determined using denaturing HPLC such as described in Nairz K et al (2002) *Proc. Natl. Acad. Sci. (U.S.A.)* 99:10575-80, and the Transgenomic WAVE® System (Transgenomic, Inc. Omaha, NE).

[0070] Oliphant et al. report a method that utilizes BeadArray™ Technology that can be used in the methods of the present invention to determine the nucleotide occurrence of a SNP (supplement to *Biotechniques*, June 2002). Additionally, nucleotide occurrences for SNPs can be determined using a DNAMassARRAY system (SEQUENOM, San Diego,

CA). This system combines proprietary SpectroChips™, microfluidics, nanodispensing, biochemistry, and MALDI-TOF MS (matrix-assisted laser desorption ionization time of flight mass spectrometry).

[0071] As another example, the nucleotide occurrences of bovine SNPs in a sample can be determined using the SNP-IT™ method (Beckman Coulter, Fullerton, CA). In general, SNP-IT™ is a 3-step primer extension reaction. In the first step a target polynucleotide is isolated from a sample by hybridization to a capture primer, which provides a first level of specificity. In a second step the capture primer is extended from a terminating nucleotide triphosphate at the target SNP site, which provides a second level of specificity. In a third step, the extended nucleotide trisphosphate can be detected using a variety of known formats, including: direct fluorescence, indirect fluorescence, an indirect colorimetric assay, mass spectrometry, fluorescence polarization, etc. Reactions can be processed in 384 well format in an automated format using a SNPstream™ instrument (Beckman Coulter, Fullerton, CA). Reactions can also be analyzed by binding to Luminex biospheres (Luminex Corporation, Austin, TX, Cai. H.. (2000) *Genomics* 66(2):135-43.). Other formats for SNP detection include TaqMan™ (Applied Biosystems, Foster City, CA), Rolling circle (Hatch et al (1999) *Genet. Anal.* 15: 35-40, Qi et al (2001) *Nucleic Acids Research* Vol. 29 e116), fluorescence polarization (Chen, X., et al. (1999) *Genome Research* 9:492-498), SNaPShot (Applied Biosystems, Foster City, CA) (Makridakis, N.M. et al. (2001) *Biotechniques* 31:1374-80.), oligo-ligation assay (Grossman, P.D., et al. (1994) *Nucleic Acids Research* 22:4527-4534), locked nucleic acids (LNATM, Link, Technologies LTD, Lanarkshire, Scotland, EP patent 1013661, US patent 6,268,490), Invader Assay (Aclara Biosciences, Wilkinson, D. (1999) *The Scientist* 13:16), padlock probes (Nilsson et al. *Science* (1994), 265: 2085), Sequence-tagged molecular inversion probes (similar to padlock probes) from ParAllele Bioscience (South San Francisco, CA; Hardenbol, P. et al. (2003) *Nature Biotechnology* 21:673-678), Molecular Beacons (Marras, S.A. et al. (1999) *Genet Anal.* 14:151-156), the READIT™ SNP Genotyping System from Promega (Madison, WI) (Rhodes R.B. et al. (2001) *Mol Diagn.* 6:55-61), Dynamic Allele-Specific Hybridization (DASH) (Prince, J.A. et al. (2001) *Genome Research* 11:152-162), the Qbead™ system (quantum dot encoded microspheres conjugated to allele-specific oligonucleotides)(Xu H. et al. (2003)

*Nucleic Acids Research* 31:e43), Scorpion primers (similar to molecular beacons except unimolecular) (Thelwell, N. et al. (2000) *Nucleic Acids Research* 28:3752-3761), and Magiprobe (a novel fluorescence quenching-based oligonucleotide probe carrying a fluorophore and an intercalator)(Yamane A. (2002) *Nucleic Acids Research* 30:e97). In addition, Rao, K.V.N. et al. ((2003) *Nucleic Acids Research*. 31:e66), recently reported a microsphere-based genotyping assay that detects SNPs directly from human genomic DNA. The assay involves a structure-specific cleavage reaction, which generates fluorescent signal on the surface of microspheres, followed by flow cytometry of the microspheres. With a slightly different twist on the Sequenom technology (MALDI), Sauer et al. ((2003) *Nucleic Acids Research* 31:e63) generate charge-tagged DNA (post PCR and primer extension), using a photocleavable linker.

[0072] Accordingly, using the methods described above, the bovine haplotype allele or the nucleotide occurrence of a bovine SNP can be identified using an amplification reaction, a primer extension reaction, or an immunoassay. The bovine haplotype allele or bovine SNP can also be identified by contacting polynucleotides in the sample or polynucleotides derived from the sample, with a specific binding pair member that selectively hybridizes to a polynucleotide region comprising the bovine SNP, under conditions wherein the binding pair member specifically binds at or near the bovine SNP. The specific binding pair member can be an antibody or a polynucleotide.

[0073] The nucleotide occurrence of a SNP can be identified by other methodologies as well as those discussed above. For example, the identification can use microarray technology, which can be performed with PCR, for example using Affymetrix technologies and GenFlex Tag arrays (See e.g., Fan et al (2000) *Genome Res.* 10:853-860), or using a bovine gene chip containing proprietary SNP oligonucleotides (See e.g., Chee et al (1996), *Science* 274:610-614; and Kennedy et al. (2003) *Nature Biotech* 21:1233-1237) or without PCR, or sequencing methods such as mass spectrometry, scanning electron microscopy, or methods in which a polynucleotide flows past a sorting device that can detect the sequence of the polynucleotide. The occurrence of a SNP can be identified using electrochemical detection devices such as the eSensor<sup>TM</sup> DNA detection system (Motorola, Inc., Yu, C.J. (2001) *J. Am Chem. Soc.* 123:11155-11161). Other formats include melting curve analysis using fluorescently labeled hybridization

probes, or intercalating dyes (Lohmann, S. (2000) *Biochemica* 4, 23-28, Herrmann, M. (2000) *Clinical Chemistry* 46: 425).

[0074] The SNP detection systems of the present invention typically utilize selective hybridization. As used herein, the term "selective hybridization" or "selectively hybridize," refers to hybridization under moderately stringent or highly stringent conditions such that a nucleotide sequence preferentially associates with a selected nucleotide sequence over unrelated nucleotide sequences to a large enough extent to be useful in identifying a nucleotide occurrence of a SNP. It will be recognized that some amount of non-specific hybridization is unavoidable, but is acceptable provide that hybridization to a target nucleotide sequence is sufficiently selective such that it can be distinguished over the non-specific cross-hybridization, for example, at least about 2-fold more selective, generally at least about 3-fold more selective, usually at least about 5-fold more selective, and particularly at least about 10-fold more selective, as determined, for example, by an amount of labeled oligonucleotide that binds to target nucleic acid molecule as compared to a nucleic acid molecule other than the target molecule, particularly a substantially similar (i.e., homologous) nucleic acid molecule other than the target nucleic acid molecule. Conditions that allow for selective hybridization can be determined empirically, or can be estimated based, for example, on the relative GC:AT content of the hybridizing oligonucleotide and the sequence to which it is to hybridize, the length of the hybridizing oligonucleotide, and the number, if any, of mismatches between the oligonucleotide and sequence to which it is to hybridize (see, for example, Sambrook et al., "Molecular Cloning: A laboratory manual (Cold Spring Harbor Laboratory Press 1989)).

[0075] An example of progressively higher stringency conditions is as follows: 2 x SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2 x SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2 x SSC/0.1% SDS at about 42EC (moderate stringency conditions); and 0.1 x SSC at about 68EC (high stringency conditions). Washing can be carried out using only one of these conditions, e.g., high stringency conditions, or each of the conditions can be used, e.g., for 10-15 minutes each, in the order listed above, repeating any or all of the steps listed. However,



as mentioned above, optimal conditions will vary, depending on the particular hybridization reaction involved, and can be determined empirically.

[0076] The term "polynucleotide" is used broadly herein to mean a sequence of deoxyribonucleotides or ribonucleotides that are linked together by a phosphodiester bond. For convenience, the term "oligonucleotide" is used herein to refer to a polynucleotide that is used as a primer or a probe. Generally, an oligonucleotide useful as a probe or primer that selectively hybridizes to a selected nucleotide sequence is at least about 15 nucleotides in length, usually at least about 18 nucleotides, and particularly about 21 nucleotides or more in length.

[0077] A polynucleotide can be RNA or can be DNA, which can be a gene or a portion thereof, a cDNA, a synthetic polydeoxyribonucleic acid sequence, or the like, and can be single stranded or double stranded, as well as a DNA/RNA hybrid. In various embodiments, a polynucleotide, including an oligonucleotide (e.g., a probe or a primer) can contain nucleoside or nucleotide analogs, or a backbone bond other than a phosphodiester bond. In general, the nucleotides comprising a polynucleotide are naturally occurring deoxyribonucleotides, such as adenine, cytosine, guanine or thymine linked to 2'-deoxyribose, or ribonucleotides such as adenine, cytosine, guanine or uracil linked to ribose. However, a polynucleotide or oligonucleotide also can contain nucleotide analogs, including non-naturally occurring synthetic nucleotides or modified naturally occurring nucleotides. Such nucleotide analogs are well known in the art and commercially available, as are polynucleotides containing such nucleotide analogs (Lin et al., *Nucleic Acids Research* (1994) 22:5220-5234 Jellinek et al., *Biochemistry* (1995) 34:11363-11372; Pagratis et al., *Nature Biotechnol.* (1997) 15:68-73, each of which is incorporated herein by reference). Primers and probes can also be comprised of peptide nucleic acids (PNA) (Nielsen PE and Egholm M. (1999) *Curr. Issues Mol. Biol.* 1:89-104).

[0078] The covalent bond linking the nucleotides of a polynucleotide generally is a phosphodiester bond. However, the covalent bond also can be any of numerous other bonds, including a thiodiester bond, a phosphorothioate bond, a peptide-like bond or any other bond known to those in the art as useful for linking nucleotides to produce synthetic

polynucleotides (see, for example, Tam et al., *Nucl. Acids Res.* (1994) 22:977-986, Ecker and Crooke, *BioTechnology* (1995) 13:351360, each of which is incorporated herein by reference). The incorporation of non-naturally occurring nucleotide analogs or bonds linking the nucleotides or analogs can be particularly useful where the polynucleotide is to be exposed to an environment that can contain a nucleolytic activity, including, for example, a tissue culture medium or upon administration to a living subject, since the modified polynucleotides can be less susceptible to degradation.

[0079] A polynucleotide or oligonucleotide comprising naturally occurring nucleotides and phosphodiester bonds can be chemically synthesized or can be produced using recombinant DNA methods, using an appropriate polynucleotide as a template. In comparison, a polynucleotide or oligonucleotide comprising nucleotide analogs or covalent bonds other than phosphodiester bonds generally are chemically synthesized, although an enzyme such as T7 polymerase can incorporate certain types of nucleotide analogs into a polynucleotide and, therefore, can be used to produce such a polynucleotide recombinantly from an appropriate template (Jellinek et al., *supra*, 1995). Thus, the term polynucleotide as used herein includes naturally occurring nucleic acid molecules, which can be isolated from a cell, as well as synthetic molecules, which can be prepared, for example, by methods of chemical synthesis or by enzymatic methods such as by the polymerase chain reaction (PCR).

[0080] In various embodiments for identifying nucleotide occurrences of SNPs, it can be useful to detectably label a polynucleotide or oligonucleotide. Detectable labeling of a polynucleotide or oligonucleotide is well known in the art. Particular non-limiting examples of detectable labels include chemiluminescent labels, fluorescent labels, radiolabels, enzymes, haptens, or even unique oligonucleotide sequences.

[0081] A method of the identifying a SNP also can be performed using a specific binding pair member. As used herein, the term "specific binding pair member" refers to a molecule that specifically binds or selectively hybridizes to another member of a specific binding pair. Specific binding pair member include, for example, probes, primers, polynucleotides, antibodies, etc. For example, a specific binding pair member includes a primer or a probe that selectively hybridizes to a target polynucleotide that includes a

SNP loci or that hybridizes to an amplification product generated using the target polynucleotide as a template.

[0082] As used herein, the term "specific interaction," or "specifically binds" or the like means that two molecules form a complex that is relatively stable under physiologic conditions. The term is used herein in reference to various interactions, including, for example, the interaction of an antibody that binds a polynucleotide that includes a SNP site; or the interaction of an antibody that binds a polypeptide that includes an amino acid that is encoded by a codon that includes a SNP site. According to methods of the invention, an antibody can selectively bind to a polypeptide that includes a particular amino acid encoded by a codon that includes a SNP site. Alternatively, an antibody may preferentially bind a particular modified nucleotide that is incorporated into a SNP site for only certain nucleotide occurrences at the SNP site, for example using a primer extension assay.

[0083] A specific interaction can be characterized by a dissociation constant of at least about  $1 \times 10^{-6}$  M, generally at least about  $1 \times 10^{-7}$  M, usually at least about  $1 \times 10^{-8}$  M, and particularly at least about  $1 \times 10^{-9}$  M or  $1 \times 10^{-10}$  M or less. A specific interaction generally is stable under physiological conditions, including, for example, conditions that occur in a living individual such as a human or other vertebrate or invertebrate, as well as conditions that occur in a cell culture such as used for maintaining mammalian cells or cells from another vertebrate organism or an invertebrate organism. Methods for determining whether two molecules interact specifically are well known and include, for example, equilibrium dialysis, surface plasmon resonance, and the like.

[0084] The invention also relates to kits, which can be used, for example, to perform a method of the invention. Thus, in one embodiment, the invention provides a kit for identifying nucleotide occurrences or haplotype alleles of bovine SNPs. Such a kit can contain, for example, an oligonucleotide probe, primer, or primer pair, or combinations thereof for identifying the nucleotide occurrence of at least one bovine single nucleotide polymorphism (SNP) associated with breed, such as a SNP corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895. Such oligonucleotides being useful, for example, to identify a SNP or haplotype allele as

disclosed herein; or can contain one or more polynucleotides corresponding to a portion of a bovine gene containing one or more nucleotide occurrences associated with a bovine trait, such polynucleotide being useful, for example, as a standard (control) that can be examined in parallel with a test sample. In addition, a kit of the invention can contain, for example, reagents for performing a method of the invention, including, for example, one or more detectable labels, which can be used to label a probe or primer or can be incorporated into a product generated using the probe or primer (e.g., an amplification product); one or more polymerases, which can be useful for a method that includes a primer extension or amplification procedure, or other enzyme or enzymes (e.g., a ligase or an endonuclease), which can be useful for performing an oligonucleotide ligation assay or a mismatch cleavage assay; and/or one or more buffers or other reagents that are necessary to or can facilitate performing a method of the invention. The primers or probes can be included in a kit in a labeled form, for example with a label such as biotin or an antibody. In one embodiment, a kit of the invention provides a plurality of oligonucleotides of the invention, including one or more oligonucleotide probes or one or more primers, including forward and/or reverse primers, or a combination of such probes and primers or primer pairs. Such a kit also can contain probes and/or primers that conveniently allow a method of the invention to be performed in a multiplex format.

[0085] The kit can also include instructions for using the probes or primers to determine a nucleotide occurrence of at least one bovine SNPs.

[0086] Population-specific alleles can be used to assign a bovine animal to a particular breed. These population specific alleles are fixed in the population of interest and absent in comparison populations. The absence of an allele in a sample of individuals from any one population may be because those alleles are truly population-specific or because the frequency of those alleles is low and the sample obtained from any given population was small (Taylor, J.F., Patent: PCT/US01/47521). For admixed populations, population-specific alleles rarely occur, however the difference in allele frequency between populations may still enable their use to infer assignment of individual bovines based to a breed, these are known as population associated alleles (Kumar, P., *Heredity* 91: 43-50

(2003)). Both population specific alleles and population-associated alleles are herein referred to as Breed-Specific Markers.

[0087] In the present invention, a marker is breed specific if it has a different allele frequency in one breed relative to one or more other breeds. A similar logic was employed by Kumar, P. (*Heredity* 91: 43-50 (2003)) to genetically distinguish cattle from European *Bos taurus* breeds and Indian *Bos indicus* breeds of cattle.

[0088] In the present invention there are 4868 breed-specific SNP markers. One or more of these markers could be used to determine breed specificity and/or to assign an individual to one or more breeds with an associated probability. These markers could be used alone or in any combination.

[0089] There are two broad classes of clustering methods that are used to assign individuals to populations (Pritchard, J.K., et al., *Genetics* 155: 945-959 (2000)). These are: 1) *Distance-based methods*: These calculate a pairwise distance matrix, whose entries give the distance between every pair of individuals. 2) *Model-based methods*: These proceed by assuming that observations from each cluster are random draws from some parametric model. Inference for the parameters corresponding to each cluster is then done jointly with inference for the cluster membership of each individual, using standard statistical methods. The present disclosure includes the use of all standard statistical methods including maximum likelihood, bootstrapping methodologies, Bayesian methods and any other statistical methodology that can be employed to analyze such genome data. These statistical techniques are well known to those in the art.

[0090] Many software programs for molecular population genetics studies have been developed, their advantage lies in their pre-programmed complex mathematical techniques and ability to handle large volumes of data. Popular programs used by those in the field include, but are not limited to: TFPGA, Arlequin, GDA, GENEPOP, GeneStrut, POPGENE (Labate, J.A., *Crop Sci.* 40: 1521-1528. (2000)) and Structure. The present disclosure incorporates the use of all of the software disclosed above used to classify bovines into populations based on DNA polymorphisms as well as other software known in the art.

[0091] *Structure* has been used to determine population structure and infer assignment of individual animals to populations for livestock species including poultry (Rosenberg, N.A., et al., *Genetics*. 159: 699-713 (2001)) and bovines from South Asia (Kumar, P., *Heredity* 91: 43-50 (2003)).

[0092] As used herein, the term "hypermutable" refers to a nucleic acid sequence that is susceptible to instability, thus resulting in nucleic acid alterations. Such alterations include the deletion and addition of nucleotides. The hypermutable sequences of the invention are most often microsatellite DNA sequences which, by definition, are small tandem repeat DNA sequences. Thus, a combination of SNP analysis and microsatellite analysis may be used to infer a breed of a bovine subject.

[0093] The following examples are intended to illustrate but not limit the invention.

#### **EXAMPLE 1**

##### **IDENTIFICATION OF SNPs THAT CAN BE USED TO INFER BREED**

[0094] This example illustrates the use of whole genome shotgun sequencing and genotyping analysis to identify SNPs that can be used to infer breed of a bovine subject. Genomic DNA libraries were constructed from six breeds of cattle (Angus, Limousin, Brahman, Simmental, Holstein and Charolais). The libraries are a mixture of 2.5 and 10 kilobase inserts. Libraries from four of the cattle breeds (Angus, Brahman, Simmental and Limousin) were sequenced using the whole genome shotgun sequencing method developed by Celera Genomics (Venter, J.C. et al. (2001) *Science* 291:1304), to a depth of coverage sufficient to generate putative SNPs. The distribution of sequence fragments obtained from the cattle breeds was 32% Angus, 35% Limousin, 33% Brahman, and less than 1% Simmental. Upon whole genome assembly of the sequenced fragments, contigs were formed from consensus sequence, and sequence variants were identified and cataloged. 786,777 sequence variants that differed by a single nucleotide became candidate SNP markers for the high-density SNP map. The relative position of each candidate SNP within the bovine genome was determined using the assembled human genome as scaffolding, creating a candidate map of 242,181 human-mapped markers. Individual markers were tested to determine breed specificity within the cattle population using 196 animals representing diverse breeds (Angus, Charolais, Limousin, Hereford,

Brahman, Simmental and Gelbvieh) and to ensure Mendelian segregation (20 trios of Beefmaster parents and progeny). The SNP detection platform used was the SNP-IT™ system (Beckman Coulter, Fullerton, CA). In general, SNP-IT™ is a 3-step primer extension reaction. In the first step a target polynucleotide is isolated from a sample by hybridization to a capture primer, which provides a first level of specificity. In a second step the capture primer is extended from a terminating nucleotide triphosphate at the target SNP site, which provides a second level of specificity. In a third step, the extended nucleotide triphosphate can be detected using a variety of known formats, including, for example: direct fluorescence, indirect fluorescence, an indirect colorimetric assay, mass spectrometry, and fluorescence polarization. Reactions were processed in an automated 384 well format using a SNPstream™ instrument (Beckman Coulter, Fullerton, CA).

[0095] Allele frequencies within breed were determined using simple counting methods. Four thousand eight hundred and seventy seven breed specific markers (See SEQ ID NOS:1 to 4868 (Tables 5 and 6, filed herewith on a compact disc) and SEQ ID NOS:64887 to 64895 (Tables 3A and 3B)) were identified by analysis of over 19,000 markers tested. Details of the SNP markers that were identified as informative for breed, are listed in Tables 5 and 6 (filed herewith on a compact disc) and Tables 3A and 3B. Tables 3A and 5 provide the names of the identified SNP markers, breed specific alleles, their associated breed, allele frequencies, and breeds for which the alleles are most significantly associated (referred to as "major breeds" in the table). Tables 3B and 6 provide primer sequences (See "Forward," "Reverse," and "Extension") that were used to amplify a region that includes the SNP, and amplicon sequences that indicate the nucleotide occurrences for the SNP that were identified in brackets within the sequence.

[0096] Table 1 shows the number of SNP markers that had a non-zero allele frequency in the breed of interest. The fourth column shows the ability to distinguish a subpopulation from the total population as given by Wright's  $F_{st}$  statistic. This statistic can take on a value of between 0 (meaning no ability to distinguish subpopulations [breeds] from the total population) and 1.0 meaning that different breeds can be identified perfectly. An  $F_{st}$  value of zero means that a particular breed has the same allele frequency as the entire population (Weir, B.S. Genetic Data Analysis II. Methods for Discrete Population

Genetic Data. Sinauer Associates, Sunderland, MA. p 174. (1996)). Table 2 shows the number of SNP markers that were breed specific.

[0097] Table 1: Number of SNP markers per breed that have a non-zero allele frequency and ability to differentiate breeds from the whole population. Values for Wright's  $F_{st}$  statistic are the mean of four replicates.

Breed	Number of animals from that breed included in the analysis	Number of SNP markers per breed that have a non-zero allele frequency	Wright's $F_{st}$ statistic
Angus	27	20	.478
Brahman	27	24	.553
Charolais	26	20	.435
Gelbvieh	26	19	.448
Hereford	26	10	.622
Limousin	26	20	.448
Simmental	26	20	.411
Beefmaster	20		.343

[0098] Table 2: Number of Population Specific SNP markers per breed (i.e. at non-zero allele frequency in one breed and zero allele frequency in all other breeds)

Breed	Number of Population Specific SNP markers per breed
Angus	27
Brahman	2976
Charolais	2
Gelbvieh	3
Hereford	3
Limousin	20
Simmental	4

## **EXAMPLE 2**

### **UTILIZING A SUBSET OF BREED SPECIFIC MARKERS FOR ASSIGNING ANIMALS TO THE CORRECT BREED CATEGORY**

[0100] For this example, 48 SNP markers were selected from the 4868 markers identified to evaluate their efficacy for use in a breed specific panel. These 48 markers included MMBT05243, MMBT02545, MMBT14829, MMBT11932, MMBT23373,



MMBT08423, MMBT19771, MMBT10324, MMBT01611, MMBT08985, MMBT02110, MMBT17611, MMBT09623, MMBT06416, MMBT04048, MMBT08023, MMBT11652, MMBT21228, MMBT12147, MMBT00059, MMBT21673, MMBT09985, MMBT21682, MMBT06242, MMBT10530, MMBT06978, MMBT21426, MMBT21425, MMBT03692, MMBT11974, MMBT01651, MMBT01127, MMBT08438, MMBT03603, MMBT11644, MMBT14000, MMBT12659, MMBT21705, MMBT05684, MMBT21233, MMBT09242, MMBT09142, MMBT00761, MMBT02076, MMBT05710, MMBT21665, MMBT04130, and MMBT18672 (See Tables 3A and 3B below and Tables 5 and 6 (provided on a compact disc filed herewith)). Animals were assigned to a population based on their genetic profile of these 48 SNP markers using the software *structure* (Pritchard, J.K., et al., *Genetics* 155: 945-959 (2000)). One hundred and ninety six (196) animals were genotyped in the study from six *Bos taurus* breeds (Angus, Hereford, Simmental, Charolais, Limousin and Gelbvieh) and one *Bos indicus* (American Brahman) breed of cattle. One further breed (Beefmaster, an admixture of Brahman, Hereford and Shorthorn cattle) was included in the study – but Breed Specific Markers were not chosen specifically for this breed.

Table 3A. Allele Frequencies

Marker	Breed Specific Allele	Breeds	AF1	AF2	AF3	AF4	Major Breeds
MMBT19771	C	Angus Gelbvieh Hereford	0.59	0.06	0.75		Angus, Hereford
MMBT08985	A	Angus Charolais Gelbvieh Simmental	0.74	0.05	0.04	0.2	Angus, Simmental
MMBT21228	T	Brahman Charolais Gelbvieh Limousin	0.55	0.3	0.03	0.02	Brahman, Charolais
MMBT12147	C	Brahman Charolais Gelbvieh Simmental	0.72	0.4	0.34	0.07	Brahman Charolais Gelbvieh
MMBT00059	G	Brahman Charolais Limousin Simmental	0.82	0.28	0.04	0.19	Brahman Charolais Gelbvieh
MMBT03692	G	Angus Hereford Simmental	0.11	0.38	0.07		Hereford Angus
MMBT01127	C	Angus Brahman Charolais Hereford	0.05	0.61	0.05	0.57	Brahman Hereford
MMBT14000	G	Limousin Simmental	0.24	0.11			Limousin
MMBT21665	T	Brahman Charolais Limousin Simmental	0.43	0.04	0.1	0.25	Brahman Simmental

Table 3B. Additional bovine SNPs associated with breed

Markers	Forward	SEQ ID	Reverse	SEQ ID	Extension	SEQ ID	Sequence	SEQ ID
		NOS: 648 96- 64904		NOS: 64 905- 64913		NOS: 6 4914- 64922		NOS: 64887 -64895
MMBT19771	TAAATGGA CCTTGGCT AAGTTTG	64896	AGAACCACTT GGGGTGG	64905	GGATGGCGTTC CGTCTATTFA CCCCACTCTCT GCTGTGTGTCAT G	64914	TGTAATTTGTGTAATTTTCTGCACTCTCTGGGACAAAATCTGGTAAT TGGATTTTAAACAAACAGTTTGTCACTAATGGATAGTCTTAAAC TATTTGCTTATGCTTCAGNAATTTTCAGTGGATGACAGTGTATTYA AATGTTTGCATCAGCTTGGCTCTCAATTTGATTTTAACTTTTAAATTT AAAATAGATTGCAACTCTGATATACATTTCTCTCACTAAATGGACCT TTGCTAATTTTGGAGTGTTCATCCCTACCCCACTCTCTGCTGTG TCATG [C/G] GAGCTAAATCTTGGCGGATTCAGCTCCACCCCAAGTG GTTCGTCTCCAGAGTCTTTGCTGTACTATGGAATGCTGCTTAT TAATCTTGTCTCTTTTACCCCAAGCACCTTTAACTGCTCTGACG AATCAAGTGGCTTACCTACAGCGGCAAGAGAGAGACTGCTTTC GCATCTTAACTGCAATGCTGATTTTAAATCAGATTTAATCAAGGACA GAGGAACTAAAGCTACTCCAAACAGCTCTTTTAAATAATPACAAAGCTGGA TCTTCTGATGCAAT	64887
MMBT08985	TTCTACTT CCCCCTC TATTC	64897	AAAAAGTCT TTAAAAAGA TACGGG	64906	ACGCAGTCCA CGGTGATTTGA AAGAAACCT TAGGTTGGGG C	64915	CTCTGCACTCTGACTTTCTCTCTGCTCAAGTAGTTTATAAATCAAGTAACTT TATGGAGTGTTCAGATGCACTGGGCTTCATGCCAGCGCTTTACACACC TTGCTTAATTAATCTCTCCAGCTACCCCCAGGAGTACATATCATATAC CTCCACTCTCCAGTGGAACTGAGGCTCAGAGAGAGTAAATCAACCCC CCTGAAAGTGGCGGCGGAGATATGAATCTGCTTAAACCCATGTC TCTTCTACTTCCCCCTCTATCTGAAAGAAACCTTAGGTTTGGG C [A/T] TCTTTGTTTATAGACTGAAGCCCGTATCTTTTAAAGACTT TTTTGATGTGATCATTTTAAAGTCTTGTATGATATGTTTACATAT TGTTTCTGTTAAATTTTGGTTTGGCCGCTGAGACATGTTGGATGTC AGCTCCCGGAGGAGTCAAGCTGCAACCCCTGCGAGGTGAAGTCTCA GGCGTGGACCAACAGGAGAGCCCTGAAAGCCCTTATGTAATGTGTAT ACTGCCCCAGACCTTAGATGCTTTGCAAGTTCAGATTTTATATCTTT TTGTTTCAG	64888

MMBT121228	ATGGGGTC ATTCAGTT GACTC	64898	ACACACAA TCAGAGTGA A	64907	AGATAGAGTCG ATGCCAGCTTT CAGGCCGTGGA T	64916	CTTCTTTCCCATTCACGTGAGGCACAGAGGCTTTAAGCAGCACCA CTTGTACTGAGAGGCCCAAGAGCTGAGCTGGCAGAGAGCTCCACCA TCTGTGTGTAHAAMAGTCTCTGAGGTTCAATACCACTCCGGTCTTGG CCTTTACATAGCCACTCTGCTGAGGTGAGGACCACTGAGCAGCTGGT GCTACAGAAACCAAGCCATTCTCCCATGGGTCAATCACTTGAC TCATGTGTAATCATCTGCTGTCATCTCAGGCTCTGACCAAGGCTCT T [G/T] MTAGGTTGCCAGGTGCTCTAGTCTCTGACTTCACTCTGA TTTGGTGTGTCCTTCCCTAGCTGATGCTCCCATATGATGTCSTGNG ATGAAGGGGGGTGACGGGGGAGGGGCTGTATGATTTGCGAGGCT GTGCATCTGCTCTCTCCCTGAGGTGCTGTACTTAAAGATGA CCTCGAATAAGCCCTGACACACATGATTTCCCTCCCTTCGAAAT TGCTGACTTAAGACTCTTCTCTTTGGCAAGGTGCCAGGTAGCAGGCTC AGGGCACAGATTC	64899
MMBT12147	ACTGCACA TAGCGGCC AT	64899	TCAGTACCAG TGACTGTAAT CATGA	64908	AGGGTCTCTAC GCTGACGATCT AATTACTATTA CTTATATAAAA A	64917	CATACAGCATGTGTGTCAGTCACTTTCAGTCAATATCTGACTCTGTATTC CCATGGAATGAGCCCAAGTCTCTCTGCTACGAGATTTCCCAAGCA AGAATACGTGGAGTGGTGCATTTCTCTCCCTCCAGAGATCTTCCCA CCTAGGGATTAATCACTCATCTTTACAGCTCTCTGCTGTGGCAGACAGAT ACTTTACATAGCAGCCAGTGGGAAGCCCTACATTTCCATACATATTG GTTCAAACTGCACATAGSGCCATTCTAATTACTATTACTTAATAA AAA [C/T] ACTTTTATAATTATGCTTTAATCTCCCAATCATGATTTACAG TCATCTGTACTGACGAAGAACTTAAATAATCAAGACTAGGCGCTGGAG TCTGATTTAGATTTGTACATATAAATTACAGGTTGGAGCAGACACAA GCCAGTCTGTGCTCCATATTTAATGGAAGTTGCAAGTAATCCATGCGA GAATGACCAATATTTCTCAAGTCTCTCTGGAGACTAATGCTATTTGCAAT TTGATTTAGATTTACTTAATCTAATCTAGGAGTTGCCAGTGCAAA TATATAAATGT	64899
MMBT00059	TTCTCCCC CACAGAAC CT	64900	ATTGATTAG CTCACCTGTC AAC	64909	AGGGTCTCTAC GCTGACGATAG TTGCTCTGAG AATCTAAGGA A	64918	TCCCTACCCAGTACTCAGAAATGTGACTATATCTTGGCCACAGGCTCTT AAAGATTAATGAGCTTAATAAGTCTCTTAAGTGAAGCTCTAAGCCAAA ACCTCTTAAGCGCTTATTAAGAGAGATGACACACACCCACAGGTAAACA TACAGGGAGAGGCGCGCTACGCGCAAGGAAGACCTCTCCATCAAGGA GATAAGGTTCTAGTCTGGGAACACACATGCTACTCTTCTCTGCTCT GTGGTCTTCCCCCAGACAACTTATGTTGGCTCTAGAAACCTCAAGGA A [A/G] TACTTCAAGCTGCCATATCAACCAACGAGGTTGACAGTGA GCTAATCAATGACAGGAGAGCTTCTGCTTCGATGCCGGATGACCCA CTGGGCTTCACTAAACCCAGCTCCATGTGGAACCCACTCTTCTCTGAGGC CTCCGCTCTCCCGTCCGTGAGAGCTCTTACAGCCCCCTGCAAGTACCTC CGTGAACCTTGTGCGCAACCATGATAAGGACGCGCAGGACTTGTCTCAG CAGCTCTGGCAGGGTGGGCATACCAAGGCGCTCTCAGGTCCTCTCTGGGCG AGGAGCTG	64891

[illegible]

MMBT21665	ATGGGCAA CG	64904	CATGTATGCG CATGGATGT	64913	AGAGCGAGTGA CGCATACTATT AGGCTTACTGT GGTGATGAGAC A	64922	TCCTTTTGGCATTTATTTCTTTCTTCCTCTTTTAAATAATTAATTTAT TTACTTATTTTGTCTGCTAGGTCTGCATTTGCTGCTCAGGCTTTTCCCA GTTGTGGCAAGTAGGGCTGTTCTCTGTTTTCAGGTGGGCTTCTCCTG CCGTGGCTTCTCACTGAGTGGCTTCTGTTGGAGTATGGGCCCTAGGG CACATGGGCTTTGGTAGTTACACCTGGGGCTCTAGAGCACATGCTCA ATATTGTGGCAATAGGCATPACGGATCTTCTGATCAGGGATCA AATCC(A/G)TGCTCATCACACACATTAAGCCTTAACATCCATCCGATA CATGGTGGCAATTAATTTTCTTAGTAGTGTATTTTTCACCTTAAATAA TTCTCTTGGCAAGCATTTCTTAAAGTCTTATTTTTCACCTTAAATAA ATCCAGTTGGAGCAGCATGTTATTAAGATTTCTGATGTTATATG GGCATCTTAATTTACCAAGAACTGATTTTGGCTTAGAGATGCAGA AGCTAGTAATATGATATTAATTAATCATGCTCAGTATGCTTCAGACAC AATGTGACCTGTT	64895
-----------	----------------	-------	-------------------------	-------	---	-------	---	-------

[0101] The number of populations  $K$  in the input file was eight. The value of  $K$  was chosen by sampling 100 SNP markers at random from the population and imputing a value of  $K$  from 4 to 12. The statistic  $\ln \Pr(X|K)$  is the likelihood of the data given the assumed number of populations, where  $X$  is the genotypes of the sampled individuals and  $K$  is number of inferred populations (Rosenberg, N.A., et al., *Genetics*. 159: 699-713 (2001)). Three values of  $K$ , ( $K=7$  to 9) gave the greatest likelihood of being the true number of populations in the study. The value of  $K=8$  was chosen for future studies - the same number of known breeds in the population. The simulation was run with a burn in period of 25,000 iterations and a runtime of 100,000 iterations for four replicates. Input parameters included in the model are shown in Table 4.

Table 4: Parameters used in the model

Number of SNP markers	n (number of animals)	B (number of breeds in input)	K (number of inferred populations)	Burn in period (iterations)	Run period (iterations)
48	196	8	8	20,000	100,000

[0102] The *structure* program uses a model-based clustering method that employs a Markov Chain to estimate the posterior distribution ( $q$ ) of each individual's admixture coefficient. The mean of this posterior distribution ( $\hat{q}_k^{(i)}$ ) represents the mean of the proportion of an animal  $i$ 's genome that is derived from a parental population ( $k$ ) where  $\sum_k q_k = 1$ ,  $i = 1, \dots, 196$  animals and  $k = 1, \dots, 7$  breeds. The higher this value of  $\hat{q}$ , the greater the likelihood an animal is derived from a particular parental population  $k$ .

[0103] Table 7 shows the mean probability of assignment to a cluster for each of the eight breeds in the dataset. On average, animals from each breed were assigned to their own clusters with high probability for all of the pure-breeds represented (probability of cluster assignment ranged from .802 for Charolais to .949 for Brahman). The composite breed, Beefmaster had a lower probability of assignment to its own cluster (.539) than the seven pure breeds but a significantly non-zero (.269) probability of assignment to the Brahman cluster -one of the parental breeds of Beefmaster. The results show that using genotype information alone, animals from the same breed were assigned to their

respective population clusters with a high probability. Table 8 shows the mean individual admixture coefficient ( $\hat{q}$ ) per animal as an estimate of probability each individual animal belonging to a particular cluster with  $K=8$  different clusters.

Table 7: Probability of cluster assignment with 48 Breed-Specific SNP markers. Results are the mean of four replicate runs of structure.

Breed	Cluster							
	1	2	3	4	5	6	7	8
Angus	.009	.002	.937	.007	.005	.008	.008	.023
Brahman	.005	.949	.020	.004	.003	.006	.006	.008
Charolais	.022	.016	.010	.064	.012	.802	.032	.042
Gelbvieh	.011	.003	.024	.067	.022	.023	.843	.007
Hereford	.948	.002	.004	.008	.004	.007	.022	.005
Limousin	.023	.002	.004	.017	.922	.013	.012	.006
Simmental	.008	.003	.006	.924	.024	.013	.016	.006
Beefmaster	.027	.269	.022	.039	.009	.051	.043	.539

Table 8: Mean individual admixture coefficient ( $\hat{q}$ ) per animal as an estimate of probability each individual animal belonging to a particular cluster with  $K=8$  different clusters.

Breed of animal	Inferred population							
	1	2	3	4	5	6	7	8
Angus	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.97
Angus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
Angus	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.96
Angus	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.97
Angus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
Angus	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.98
Angus	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.96
Angus	0.18	0.01	0.02	0.00	0.13	0.02	0.05	0.59
Angus	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.95
Angus	0.02	0.01	0.01	0.00	0.05	0.01	0.02	0.89
Angus	0.01	0.00	0.01	0.00	0.01	0.00	0.02	0.95
Angus	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.97
Angus	0.02	0.01	0.01	0.00	0.01	0.01	0.01	0.93
Angus	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.95

Angus	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.98
Angus	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.96
Angus	0.01	0.01	0.00	0.00	0.05	0.00	0.01	0.91
Angus	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.97
Angus	0.09	0.01	0.01	0.00	0.07	0.02	0.05	0.75
Angus	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.95
Angus	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.97
Angus	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.97
Angus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98
Beefmaster	0.02	0.01	0.24	0.42	0.16	0.01	0.13	0.03
Beefmaster	0.01	0.01	0.93	0.02	0.01	0.02	0.00	0.01
Beefmaster	0.01	0.02	0.91	0.06	0.01	0.00	0.00	0.00
Beefmaster	0.08	0.14	0.39	0.27	0.10	0.01	0.01	0.01
Beefmaster	0.03	0.24	0.06	0.63	0.02	0.01	0.01	0.00
Beefmaster	0.02	0.01	0.35	0.47	0.01	0.01	0.01	0.13
Beefmaster	0.24	0.11	0.14	0.45	0.02	0.01	0.03	0.01
Beefmaster	0.03	0.01	0.94	0.01	0.00	0.00	0.01	0.00
Beefmaster	0.07	0.01	0.81	0.02	0.04	0.01	0.01	0.03
Beefmaster	0.08	0.14	0.05	0.65	0.03	0.01	0.03	0.01
Beefmaster	0.04	0.01	0.06	0.67	0.05	0.01	0.14	0.02
Beefmaster	0.16	0.01	0.18	0.44	0.11	0.01	0.08	0.02
Beefmaster	0.01	0.00	0.98	0.00	0.00	0.00	0.00	0.00
Beefmaster	0.01	0.00	0.96	0.02	0.00	0.00	0.00	0.00
Beefmaster	0.36	0.01	0.59	0.01	0.02	0.01	0.01	0.00
Beefmaster	0.10	0.12	0.01	0.56	0.13	0.02	0.06	0.01
Beefmaster	0.01	0.00	0.96	0.00	0.01	0.00	0.01	0.01
Beefmaster	0.20	0.06	0.06	0.52	0.09	0.03	0.03	0.01
Beefmaster	0.17	0.05	0.24	0.43	0.03	0.01	0.02	0.05
Beefmaster	0.00	0.00	0.98	0.01	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.02	0.96	0.01	0.00	0.00	0.00
Brahman	0.01	0.00	0.01	0.97	0.00	0.00	0.00	0.00
Brahman	0.01	0.00	0.02	0.94	0.01	0.01	0.01	0.01
Brahman	0.01	0.04	0.01	0.92	0.01	0.01	0.01	0.00
Brahman	0.01	0.01	0.01	0.92	0.01	0.02	0.02	0.01
Brahman	0.01	0.01	0.01	0.97	0.01	0.00	0.01	0.00
Brahman	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00
Brahman	0.01	0.01	0.00	0.96	0.01	0.01	0.01	0.00
Brahman	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
Brahman	0.05	0.00	0.03	0.85	0.02	0.01	0.01	0.04
Brahman	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
Brahman	0.01	0.01	0.02	0.94	0.01	0.01	0.01	0.00



Brahman	0.00	0.00	0.01	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.01	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.01	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.01	0.97	0.00	0.00	0.00	0.01
Brahman	0.00	0.00	0.01	0.97	0.01	0.00	0.00	0.00
Brahman	0.00	0.01	0.00	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00
Brahman	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00
Brahman	0.02	0.00	0.03	0.89	0.02	0.01	0.02	0.01
Charolais	0.84	0.01	0.02	0.02	0.03	0.01	0.06	0.03
Charolais	0.38	0.01	0.15	0.01	0.13	0.00	0.13	0.19
Charolais	0.03	0.02	0.89	0.00	0.02	0.01	0.02	0.01
Charolais	0.72	0.00	0.01	0.01	0.24	0.01	0.01	0.00
Charolais	0.74	0.01	0.16	0.01	0.02	0.04	0.01	0.01
Charolais	0.97	0.00	0.01	0.00	0.01	0.00	0.00	0.00
Charolais	0.96	0.01	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.96	0.00	0.01	0.00	0.02	0.00	0.01	0.00
Charolais	0.97	0.00	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.92	0.04	0.02	0.02	0.01	0.00	0.00	0.00
Charolais	0.88	0.00	0.01	0.00	0.01	0.01	0.08	0.00
Charolais	0.67	0.02	0.12	0.00	0.15	0.01	0.02	0.02
Charolais	0.96	0.00	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.97	0.00	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.68	0.01	0.02	0.00	0.03	0.01	0.25	0.00
Charolais	0.96	0.01	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.14	0.01	0.01	0.00	0.83	0.01	0.01	0.00
Charolais	0.96	0.00	0.00	0.00	0.01	0.01	0.02	0.00
Charolais	0.97	0.00	0.00	0.00	0.01	0.00	0.01	0.00
Charolais	0.95	0.01	0.00	0.01	0.01	0.01	0.01	0.00
Charolais	0.95	0.01	0.01	0.00	0.01	0.00	0.01	0.00
Charolais	0.96	0.01	0.01	0.00	0.01	0.00	0.01	0.00
Charolais	0.94	0.01	0.01	0.00	0.02	0.01	0.02	0.00
Charolais	0.90	0.05	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.96	0.01	0.01	0.00	0.01	0.01	0.01	0.00
Charolais	0.95	0.00	0.01	0.00	0.03	0.01	0.01	0.00
Gelbvieh	0.05	0.01	0.01	0.00	0.93	0.00	0.01	0.00
Gelbvieh	0.01	0.00	0.01	0.05	0.77	0.01	0.13	0.03
Gelbvieh	0.01	0.00	0.00	0.00	0.96	0.01	0.01	0.00

Gelbvieh	0.01	0.01	0.00	0.00	0.97	0.00	0.01	0.00
Gelbvieh	0.01	0.01	0.00	0.00	0.96	0.01	0.01	0.00
Gelbvieh	0.01	0.00	0.00	0.00	0.95	0.03	0.01	0.00
Gelbvieh	0.01	0.01	0.00	0.00	0.95	0.01	0.02	0.00
Gelbvieh	0.05	0.06	0.09	0.00	0.64	0.01	0.01	0.14
Gelbvieh	0.15	0.01	0.01	0.00	0.79	0.01	0.03	0.01
Gelbvieh	0.01	0.01	0.00	0.00	0.96	0.00	0.01	0.00
Gelbvieh	0.06	0.01	0.01	0.00	0.86	0.01	0.04	0.01
Gelbvieh	0.01	0.00	0.01	0.00	0.66	0.01	0.02	0.29
Gelbvieh	0.01	0.00	0.00	0.00	0.76	0.17	0.05	0.00
Gelbvieh	0.01	0.01	0.01	0.00	0.51	0.01	0.46	0.00
Gelbvieh	0.01	0.01	0.00	0.00	0.97	0.00	0.01	0.00
Gelbvieh	0.00	0.00	0.00	0.00	0.97	0.01	0.00	0.00
Gelbvieh	0.01	0.00	0.00	0.00	0.96	0.01	0.01	0.00
Gelbvieh	0.01	0.02	0.01	0.00	0.32	0.02	0.61	0.02
Gelbvieh	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00
Gelbvieh	0.01	0.00	0.01	0.00	0.92	0.01	0.04	0.00
Gelbvieh	0.03	0.00	0.01	0.00	0.94	0.01	0.01	0.01
Gelbvieh	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00
Gelbvieh	0.02	0.00	0.01	0.00	0.89	0.01	0.07	0.00
Gelbvieh	0.01	0.01	0.00	0.00	0.96	0.01	0.01	0.00
Gelbvieh	0.01	0.00	0.00	0.00	0.88	0.06	0.05	0.00
Hereford	0.01	0.95	0.01	0.00	0.01	0.01	0.01	0.01
Hereford	0.01	0.98	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.01	0.97	0.00	0.00	0.01	0.00	0.00	0.00
Hereford	0.01	0.97	0.01	0.00	0.01	0.00	0.01	0.00
Hereford	0.01	0.93	0.02	0.00	0.01	0.00	0.01	0.04
Hereford	0.02	0.94	0.01	0.00	0.01	0.01	0.01	0.01
Hereford	0.01	0.97	0.00	0.00	0.01	0.00	0.01	0.00
Hereford	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.09	0.77	0.01	0.00	0.07	0.02	0.04	0.01
Hereford	0.01	0.98	0.00	0.00	0.01	0.00	0.00	0.00
Hereford	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.01	0.97	0.00	0.00	0.01	0.00	0.00	0.00
Hereford	0.41	0.16	0.03	0.00	0.36	0.01	0.03	0.01
Hereford	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.01	0.98	0.00	0.00	0.01	0.00	0.00	0.00
Hereford	0.12	0.74	0.01	0.00	0.08	0.02	0.03	0.01
Hereford	0.01	0.98	0.00	0.00	0.01	0.00	0.00	0.00
Hereford	0.01	0.97	0.00	0.00	0.01	0.00	0.00	0.00
Hereford	0.00	0.95	0.00	0.00	0.00	0.01	0.00	0.03

Hereford	0.01	0.97	0.00	0.00	0.01	0.00	0.01	0.00
Hereford	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.03	0.91	0.01	0.00	0.02	0.01	0.01	0.01
Hereford	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
Hereford	0.01	0.96	0.01	0.00	0.01	0.01	0.01	0.00
Hereford	0.01	0.97	0.01	0.00	0.01	0.00	0.01	0.00
Limousin	0.01	0.01	0.01	0.00	0.01	0.97	0.01	0.00
Limousin	0.01	0.01	0.01	0.00	0.03	0.92	0.02	0.00
Limousin	0.01	0.00	0.01	0.00	0.01	0.97	0.01	0.00
Limousin	0.02	0.00	0.01	0.00	0.07	0.89	0.01	0.01
Limousin	0.01	0.00	0.00	0.00	0.01	0.97	0.01	0.00
Limousin	0.01	0.00	0.00	0.00	0.00	0.98	0.00	0.00
Limousin	0.01	0.00	0.00	0.00	0.01	0.95	0.02	0.00
Limousin	0.01	0.01	0.00	0.00	0.01	0.96	0.01	0.00
Limousin	0.02	0.01	0.01	0.00	0.01	0.94	0.01	0.02
Limousin	0.01	0.00	0.00	0.00	0.01	0.97	0.01	0.00
Limousin	0.02	0.02	0.02	0.00	0.05	0.86	0.04	0.01
Limousin	0.01	0.00	0.00	0.00	0.00	0.97	0.01	0.01
Limousin	0.01	0.00	0.00	0.00	0.01	0.97	0.01	0.00
Limousin	0.09	0.05	0.01	0.00	0.18	0.56	0.10	0.01
Limousin	0.29	0.01	0.01	0.00	0.01	0.67	0.01	0.00
Limousin	0.01	0.00	0.00	0.00	0.01	0.96	0.01	0.00
Limousin	0.00	0.00	0.00	0.00	0.01	0.97	0.01	0.01
Limousin	0.00	0.00	0.00	0.00	0.00	0.98	0.01	0.00
Limousin	0.14	0.00	0.02	0.01	0.01	0.80	0.03	0.00
Limousin	0.04	0.00	0.00	0.00	0.01	0.90	0.05	0.00
Limousin	0.00	0.00	0.00	0.00	0.00	0.98	0.00	0.00
Limousin	0.01	0.00	0.00	0.00	0.01	0.97	0.01	0.00
Limousin	0.00	0.00	0.03	0.00	0.00	0.93	0.03	0.00
Limousin	0.03	0.01	0.01	0.00	0.02	0.92	0.01	0.00
Limousin	0.01	0.01	0.09	0.00	0.01	0.69	0.18	0.00
Limousin	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00
Simmental	0.35	0.02	0.01	0.00	0.15	0.01	0.45	0.01
Simmental	0.01	0.00	0.00	0.00	0.00	0.01	0.97	0.00
Simmental	0.01	0.00	0.00	0.00	0.01	0.00	0.96	0.01
Simmental	0.02	0.01	0.01	0.00	0.01	0.02	0.94	0.01
Simmental	0.22	0.01	0.01	0.00	0.09	0.01	0.66	0.01
Simmental	0.01	0.00	0.00	0.00	0.01	0.01	0.97	0.00
Simmental	0.02	0.00	0.01	0.00	0.49	0.00	0.48	0.00
Simmental	0.01	0.00	0.00	0.00	0.01	0.01	0.96	0.01
Simmental	0.02	0.01	0.00	0.00	0.02	0.01	0.90	0.04
Simmental	0.01	0.00	0.01	0.00	0.01	0.01	0.95	0.01

Simmental	0.02	0.00	0.01	0.01	0.01	0.01	0.95	0.00
Simmental	0.00	0.00	0.00	0.00	0.01	0.00	0.98	0.00
Simmental	0.00	0.11	0.00	0.01	0.01	0.00	0.87	0.00
Simmental	0.02	0.01	0.01	0.00	0.02	0.37	0.57	0.01
Simmental	0.01	0.00	0.01	0.00	0.00	0.01	0.97	0.01
Simmental	0.02	0.00	0.00	0.00	0.01	0.06	0.90	0.00
Simmental	0.01	0.00	0.03	0.00	0.02	0.01	0.93	0.00
Simmental	0.01	0.00	0.00	0.00	0.03	0.00	0.93	0.02
Simmental	0.05	0.02	0.01	0.04	0.01	0.01	0.86	0.01
Simmental	0.03	0.01	0.03	0.00	0.02	0.01	0.91	0.00
Simmental	0.01	0.00	0.00	0.00	0.01	0.00	0.97	0.00
Simmental	0.01	0.00	0.00	0.00	0.01	0.01	0.96	0.00
Simmental	0.02	0.00	0.00	0.00	0.10	0.38	0.48	0.00
Simmental	0.01	0.01	0.01	0.00	0.01	0.04	0.92	0.00

**EXAMPLE 3****IDENTIFICATION OF ANGUS VS NON-ANGUS**

[0104] In another aspect of the invention, 12 SNP markers were selected and tested for their ability to distinguish Angus from non-Angus breeds. These markers included MMBT05243, MMBT02545, MMBT14829, MMBT11932, MMBT23373, MBT08423, MMBT19771, MMBT10324, MMBT01611, MMBT08985, MMBT02110, and MMBT17611 (See Tables 5 and 6). Parameters used in this analysis are shown in Table 9.

Table 9. Parameters used in the model

Number of SNP markers	n (number of animals)	B (number of breeds in input)	K (number of inferred populations)	Burn in period (iterations)	Run period (iterations)
12	196	8	2 (Angus and non-Angus)	50,000	250,000

[0105] The mean individual admixture coefficient ( $\hat{q}$ ) for Angus and non-Angus animals was .968 and .924, respectively, indicating a very high probability of assigning an animal from one of these populations to the correct cluster. The mean individual

admixture coefficient ( $\hat{q}$ ) per animal is shown in Table 10.

Table 10: Mean individual admixture coefficient ( $\hat{q}$ ) per animal as an estimate of probability each individual animal belongs to a particular cluster.

Breed of animal	Inferred population	
	1	2
Angus	0.99	0.02
Angus	1.00	0.00
Angus	0.98	0.02
Angus	0.99	0.01
Angus	1.00	0.01
Angus	0.99	0.01
Angus	0.99	0.01
Angus	0.40	0.60
Angus	0.98	0.02
Angus	0.88	0.12
Angus	0.98	0.02
Angus	0.98	0.02
Angus	0.95	0.05
Angus	0.99	0.01
Angus	0.99	0.01
Angus	0.97	0.03
Angus	0.95	0.05
Angus	0.99	0.01
Angus	0.46	0.54
Angus	0.95	0.05
Angus	0.97	0.03
Angus	0.98	0.02
Angus	0.99	0.01
non Angus	0.02	0.98
non Angus	0.02	0.99
non Angus	0.02	0.98
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.01	0.99
non Angus	0.02	0.99
non Angus	0.01	0.99
non Angus	0.08	0.92
non Angus	0.31	0.69

non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.02	0.98
non Angus	0.04	0.97
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.04	0.96
non Angus	0.17	0.83
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.04	0.96
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.04	0.96
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.58	0.42
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00

non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.33	0.67
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.34	0.66
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.05	0.95
non Angus	0.01	0.99
non Angus	0.02	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.02	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00

non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.16	0.85
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.17	0.83
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.08	0.92
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.07	0.93
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.40	0.60
non Angus	0.02	0.98
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99



non Angus	0.26	0.74
non Angus	0.10	0.91
non Angus	0.02	0.98
non Angus	0.01	0.99
non Angus	0.26	0.74
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.42	0.59
non Angus	0.01	0.99
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.01	1.00
non Angus	0.09	0.91
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.28	0.73
non Angus	0.01	0.99
non Angus	0.01	0.99
non Angus	0.10	0.90
non Angus	0.01	0.99
non Angus	0.05	0.95
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.02	0.98
non Angus	0.01	1.00
non Angus	0.01	0.99
non Angus	0.02	0.98
non Angus	0.01	0.99
non Angus	0.05	0.96
non Angus	0.03	0.98

[0106] This tool can be used to identify the percentage of each breed comprising an individual animal. These data can be used to manage feedlot cattle for specific growth

and development traits. In one aspect, cattle identified from the exotic breed type (Charolais, Gelbvieh, Limousin, and Simmental) have high growth rate and large harvest sizes. These cattle can be managed to maximize growth rate and lean meat yield. Cattle identified as English type (Angus and Hereford) produce high quality meat products and can be marketed into the high quality yield grade. Breed specified products can certify that the product is in fact from the breed labeled.

[0107] Although the invention has been described with reference to the above example, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. A method to infer breed of a bovine subject from a nucleic acid sample of the bovine subject, comprising identifying in the nucleic acid sample, at least one nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with a breed, thereby inferring the breed of the bovine subject.
2. The method of claim 1, wherein the nucleotide occurrence of at least 2 of the SNPs is determined to infer the breed.
3. The method of claim 2, wherein the nucleotide occurrences of the at least 2 SNPs comprise a haplotype allele, and wherein the method infers the breed using the identified haplotype allele.
4. The method of claim 3, further comprising identifying a diploid pair of haplotype alleles and inferring the breed using the diploid pair of haplotype alleles.
5. The method of claim 1, wherein the identified nucleotide occurrence is associated with Angus, Holstein, Limousin, Brahman, Simmental, Hereford, Gelbvieh or Charolais cattle.
6. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Angus cattle is identified.
7. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Brahman cattle is identified.
8. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Limousin cattle is identified.

9. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Simmental cattle is identified.

10. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Hereford cattle is identified.

11. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Charolais cattle is identified.

12. The method of claim 5, wherein at least one nucleotide occurrence of at least one SNP listed in Table 5 or Table 3A as associated with Gelbvieh cattle is identified.

13. The method of claim 1, further comprising marketing the subject prior to harvest based on the inferred breed.

14. The method of claim 1, further comprising managing the subject in the feedlot to obtain improved performance based on known characteristics of the inferred breed for the subject.

15. The method of claim 1, further comprising identifying a nucleotide sequence of a hypermutable sequence in the sample, and inferring breed based on the at least one nucleotide occurrence and the nucleotide sequence of the hypermutable sequence.

16. The method of claim 1, further comprising marketing a product of the bovine subject based on the inferred breed.

17. The method of claim 16, wherein the marketing of the product comprises marketing the product under a trademark specific for a breed.

18. A method for determining a nucleotide occurrence of a single nucleotide polymorphism (SNP) in a bovine sample, comprising:

- a) contacting a bovine polynucleotide in the sample with an oligonucleotide that binds to a target region of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895; and
- b) determining the nucleotide occurrence of a single nucleotide polymorphism (SNP) corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the determination comprises analyzing binding of the oligonucleotide, or detecting an amplification product generated using the oligonucleotide, thereby determining the nucleotide occurrence of the SNP.

19. The method of claim 18, wherein the oligonucleotide binds to a target sequence that comprises one of the SNPs, and the nucleotide occurrence is determined based on the binding of the oligonucleotide to the target sequence.

20. The method of claim 18, wherein the bovine polynucleotide is contacted with a pair of oligonucleotides that comprise a primer pair, and the nucleotide occurrence is determined using an amplification product generated using the primer pair.

21. The method of claim 20, wherein at least one primer of the primer pair is any one of SEQ ID NOS:Forward1 to Reverse4868.

22. The method of claim 21, wherein the primer pair is any of the forward and reverse primer pairs listed in Table 6.

23. The method of claim 19, wherein the terminal nucleotide of the oligonucleotide binds to the SNP.

24. The method of claim 23, wherein the method comprises detecting an extension product generated using the oligonucleotide as a primer.

25. The method of claim 20, wherein the terminal nucleotide of each oligonucleotide of a pair of oligonucleotides is complementary to a different nucleotide at position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, or a complement thereof.

26. The method of claim 22, wherein the oligonucleotide is any one of SEQ ID NO:4879 to 9736 or SEQ ID NO:64914 to 64922.

27. The method of claim 18, wherein the polynucleotide is contacted with a pair of oligonucleotides each comprising a different detectable label.

28. An isolated polynucleotide comprising at least 50 contiguous nucleotides of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the polynucleotide is less than or equal to about 500,000 nucleotides in length.

29. The isolated polynucleotide of claim 28, wherein the polynucleotide comprises 100 contiguous nucleotides of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

30. The isolated polynucleotide of claim 28, wherein the polynucleotide comprises any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

31. The isolated polynucleotide of claim 29, wherein the polynucleotide further comprises a detectable label at a position corresponding to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

32. An isolated oligonucleotide that binds to any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the oligonucleotide is at least 10 nucleotides in length.

33. The oligonucleotide of claim 32, wherein the oligonucleotide is any one of SEQ ID NOS:4869 to 19472 or SEQ ID NOS:64896 to 64922.

34. The isolated oligonucleotide of claim 32, wherein the oligonucleotide is at least 15 nucleotides in length

35. A primer pair that binds to a first target region and a second target region of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the a first primer of the primer pair and a second primer of the primer pair are at least 10 nucleotides in length and bind opposite strands of the target region, and prime polynucleotide synthesis from the target region in opposite directions across position 300 of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

36. The primer pair of claim 35, wherein the primer is a forward primer and a reverse primer listed in Table 6.

37. An isolated oligonucleotide comprising 10 nucleotides, that selectively binds to a target polynucleotide of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein a terminal nucleotide of the isolated oligonucleotide binds to position 299, 300, or 301 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

38. The isolated oligonucleotide of claim 37, wherein the oligonucleotide is at least 15 nucleotides in length.

39. The isolated oligonucleotide of claim 37, wherein the terminal nucleotide binds to position 300 of any one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895.

40. The isolated oligonucleotide of claim 37, wherein the oligonucleotide is any one of SEQ ID NO:4879 to 9736 or SEQ ID NO:64914 to 64922.

41. An isolated vector comprising a polynucleotide of claim 28.

42. An isolated cell comprising the vector of claim 41.

43. A method for sorting bovine subjects, comprising:

a) inferring a breed for a first bovine subject from a nucleic acid sample of the first bovine subject, by a method comprising identifying a nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of at least one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with the breed; and

b) sorting the first bovine subject based on the inferred breed, and repeating for additional subjects, thereby sorting bovine subjects.

44. The method of claim 43, wherein the bovine subjects are sorted based on whether they are Angus, Limousin, Brahman, Simmental, Hereford, Gelbvieh or Charolais cattle.

45. A method for breeding a bovine subject, comprising:

a) inferring breed of a bovine candidate for use in breeding programs from a nucleic acid sample of the bovine candidate by a method comprising identifying the nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of at least one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with breed; and

b) determining whether to select the individual for use in breeding programs based on the inferred breed, thereby breeding the bovine subject.



46. The method of claim 45, wherein the selection is implemented at the elite or breeding nucleus level or at the multiplier or foundation animal level.

47. The method of claim 45, wherein the selected bovine subject is used for breeding.

48. A bovine subject resulting from the selection and breeding according to claim 45.

49. The method of claim 48, wherein the selected bovine subject is used for cloning.

50. A method of providing labeling accuracy for breed identified meat products, comprising:

a) inferring breed of a bovine candidate for use in branded meat products from a nucleic acid sample of the bovine candidate by a method comprising identifying the nucleotide occurrence of at least one single nucleotide polymorphism (SNP) corresponding to position 300 of at least one of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with breed; and

b) determining whether to brand the harvested product based on the inferred breed, thereby providing labeling accuracy for breed identified meat products.

51. The method of claim 50, wherein the inferred breed is Angus, Limousin, Brahman, Simmental, Hereford, Gelbvieh or Charolais.

52. A kit for determining nucleotide occurrences or haplotype alleles of bovine SNPs, comprising an oligonucleotide probe, primer, or primer pair, or combinations thereof, for identifying the nucleotide occurrence of at least one bovine single nucleotide polymorphism (SNP) corresponding to position 300 of SEQ ID NOS:1 to 4868 or SEQ ID NOS:64887 to 64895, wherein the SNP is associated with breed.

53. The kit of claim 52, further comprising one or more detectable labels.

54. The kit of claim 52, wherein the kit comprises a plurality of oligonucleotide probes, primers, or primer pairs, or combinations thereof, for identifying the nucleotide occurrence of at least two of the SNPs.

55. The kit of claim 52, wherein the kit comprises at least two probes, primers, or primer pairs for identifying the nucleotide occurrence of at least two SNPs that comprise a haplotype, and wherein the kit allows a determination of a haplotype allele that is associated with the trait.